DOCUMENT RESUME

ED 316 429 SE 051 260

AUTHOR Olinick, Michael
TITLE Reasoning with Probabilities.
INSTITUTION Council of Independent Colleges, Washington, D.C.
SPONS AGENCY EXXON Education Foundation, New York, N.Y.
PUB DATE 87
NOTE 62p.
AVAILABLE FROM Council of Independent Colleges Publications Department, Technology Project, One Dupont Circle, Suite 320, Washington, DC 20036.
PUB TYPE Guides - Classroom Use - Materials (For Learner) (051)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *College Mathematics; Communications; Community Colleges; Higher Education; Mathematics Education; *Nonmajors; *Probability; *Science and Society; Secondary Education; Secondary School Mathematics; *Technological Advancement; Technology

ABSTRACT

Technological advances during the past few decades have revolutionized many complex systems that influence human activity. As the rate of technological progress accelerates, these systems will become more complex, and new ones will evolve. Citizens in a technological society need to be able to make intelligent choices about how technology will affect their lives. This requires a better understanding of the technology itself; its potential and its limitations. This is a module for a course designed to introduce students, particularly those not majoring in the sciences or mathematics, to technological concepts. Written for students with no background in probability, this module introduces the fundamental concepts through an extensive series of examples carefully selected to match student interests and to establish the foundation for later applied mathematics (e.g., queuing) or technology studies. (CW)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED316429

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY
S.G. Pelletier
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

# Reasoning with Probabilities

by Michael Olinick

**Written for students with no background in probability, this module introduces the fundamental concepts through an extensive series of examples carefully selected to match student interests and to lay the foundation for later applied-mathematics (e.g., queuing) or technology studies.**

*Designed to further student understanding of technological issues, this module is one in a series developed in a project on "Technology and the Liberal Arts," which is sponsored by CIC—the Council of Independent Colleges, Washington, D.C.*

PUBLISHED BY: CIC, THE COUNCIL OF INDEPENDENT COLLEGES

SE051260



BEST COPY AVAILABLE

Technological advances in just the past few decades have revolutionized many of the complex systems that guide our lives—the way we travel, for example, and how we communicate with one another, heal our sick, even the way we grow our food. As the rate of technological progress accelerates, these systems will become more complex, and new ones will evolve. Technology will continue to be a preeminent force in our lives.

As citizens in a technological society, therefore, we need to be able to make intelligent choices about how technology will affect our future lives. This requires that we have a better understanding of technology itself. We need to understand its potential and limitations. Further, we need to understand the very *process* of technology, and have a sense of its methodology and rules.

The Council of Independent Colleges has been deeply involved in these issues since 1984, in a project entitled "Technology and the Liberal Arts." Working with faculty from small liberal arts colleges, CIC has sought to integrate the study of technology and technological issues into the liberal arts curriculum. Our goal is to improve the understanding of technology on the part of undergraduate students—the future citizens of the technological society.

This course module is one in a series developed in the CIC project under a grant from the Exxon Education Foundation. Each module is designed to introduce students, particularly those not majoring in the sciences, to a technological issue or concept. In the process of completing the exercises in the modules, we also hope students will develop an idea of the methodology of technological inquiry. Our goal in publishing the modules is to make teaching materials available for use in courses across the liberal arts disciplines. We are deeply indebted to the Exxon Education Foundation and its executive director, Arnold R. Shore, for supporting this endeavor.

I want to also acknowledge the many college faculty who have participated in CIC's technology project. To each, I offer my sincere thanks for pioneering the important effort to bring technology into the mainstream of the curriculum.

Allen P. Splete
President
The Council of Independent Colleges
October 1987

*In addition to the Exxon Education Foundation, the following sources have provided funding for the CIC project on Technology and the Liberal Arts: Monsanto, Pew Memorial Trust, AT&T, Ford Foundation, Consortium for the Advancement of Private Higher Education, Pfizer, and Xerox.*

# Reasoning With Probabilities

by Michael Olinick

Professor, Department of Mathematics
and Computer Science
Middlebury College
Middlebury, Vermont

In the 17th Century, Isaac Newton (1642-1727) used the newly-forged tools of calculus along with a handful of simplifying physical assumptions to investigate the motions of the Earth, its moon, the other planets and the sun. Although the focus of Newton's study was the heavenly bodies, his approach could also be applied to the analysis of more familiar situations.

We can predict, for example, that if an apple initially F feet above the ground is released, it will hit the ground a certain number T of seconds later. The numbers F and T are related by the equation

$$F = 16 T^2$$

From this equation, we can make certain predictions. If we observe that an apple takes 2 seconds to fall to the earth, the equation tells us that it fell from a height of 64 feet. We can also work backwards, if desired, to compute the time it takes the object to fall from a known height.

*Copyright © 1987 Council of Independent Colleges*

### *1.2 Rising Populations*

In the classical model of exponential growth, one investigates the assumption that a population increases at a rate proportional to its numbers. If it takes D hours for the population to double in size, then it can be shown that the population P at any time t is given by the equation

$$P(t) = P_0 \, 2^{t/D}$$

where $P_0$ is the initial size of the population.

### *1.3 Deterministic Models*

These two examples—falling bodies and rising populations—are examples of deterministic models. The assumption behind a deterministic model of a process which changes over time is that the entire future behavior is exactly and explicitly determined by the present status of the system and the forces acting on it. In other words, if we know everything about the system at a particular moment, then we can at least in principle predict its behavior at every future instant.

This belief was the driving force that led to the very fruitful development of the physical sciences during the 18th and 19th centuries. Much of our understanding of the behavior of physical systems we observe in daily life comes from deterministic models. Physical scientists and engineers developed powerful analytic techniques to analyze more and more complex deterministic models. The availability of these techniques and the predictive success of these models in the physical sciences motivated many thinkers to employ similar models in the study of social and biological systems.

### *1.4 The Need for Probability Models*

There are several important objections to the use of deterministic models in the social and life sciences. In the first place, some philosophers argue that deterministic models of social systems must necessarily assume that human beings have no free will; few people are willing to accept this view of humans. Social problems deal with individuals or groups of individuals. We can never completely predict the exact future behavior of any person in a specific situation, no matter how well we understand the person or the social context.

A second objection arises from the discovery of Werner Heisenberg (1901 - 1976) in the early part of this century that purely deterministic models are

insufficient even to study physical processes. Heisenberg showed that it is impossible, even in theory, to know the exact state of a physical system. We can conclude from models only the average behavior of a group of atoms, but we cannot assert with certainty anything about the future course of an individual atom. This observation has fundamentally affected the physical sciences and led to the introduction of probabilistic models.

Nobel Laureate Richard Feynman emphasizes the new probabilistic viewpoint of physics in this excerpt from his book *Lectures on Physics:*

> We would like to emphasize a very important difference between classical and quantum mechanics. We have been talking about the probability that an electron will arrive in a given circumstance. We have implied that in our experimental arrangement (or even in the best possible one) it would be impossible to predict exactly what would happen. We can only predict the odds! This would mean, if it were true, that physics has given up on the problem of trying to predict exactly what will happen in a definite circumstance. Yes! Physics *has* given up. We do not know how to predict what would happen in a given circumstance, and we believe that it is impossible—that the only thing that can be predicted is the probability of different events. It must be recognized then that this is a retrenchment in our earlier ideal of understanding nature. It may be a backward step, but no one has seen a way to avoid it...We suspect very strongly that it is something that will be with us forever...that this is the way nature really is. (Richard P. Feynman, Robert B. Leighton and Matthew Sands, *The Feynman Lectures on Physics*, Volume 3, Reading, Massachusetts: Addison-Wesley, 1965.)

We see then that the probabilistic viewpoint will come to dominate quantitative models and mathematical thinking about physical, scientific, technological and social problems. This is a point of view in which concepts of chance and random events play an essential role. In this module we will introduce some of the elementary concepts of probability to investigate four specific problems.

# Chapter 2

## Four Problems

In this chapter we will introduce four problems we hope you'll find interesting and provocative. As we develop the tools for reasoning with probabilities, our ultimate goals include solving these problems.

### *2.1 Problem 1: Women and Witchcraft*

In 17th century America, a good many people were accused of witchcraft. Both men and women faced witchcraft charges, but not in equal numbers. Historians wish to know if there was a sex bias in these trials: were women

more likely to be convicted than men? In one set of data from New England, for example, we find 21 men and 68 women facing witchcraft charges; only 2 of the men were convicted while 14 of the women were found guilty. Were men and women judged by different standards?

### *2.2 Problem 2: Screening For AIDS*

Acquired immune deficiency syndrome (AIDS) is a fatal contagious disease which many observers believe has reached epidemic proportions in the United States. To halt the spread of the disease, we need first to identify carriers of the AIDS virus. Scientists are searching for an effective method to identify AIDS virus carriers.

One technique is a blood test which detects the presence of certain antibodies in the body. Although such a test may be extremely accurate, it is subject to two possible errors: it may fail to detect the illness in someone who actually suffers from it. Secondly, it may "detect" AIDS in a person who is free of the disease.

Suppose the blood test correctly identifies AIDS in 95 percent of the population which has the disease. Furthermore, suppose that 99 percent of those tested who are free of the AIDS virus also receive a negative test result. If your blood is tested and the result is positive, how worried should you be? What are the chances that you actually have AIDS?

### *2.3 Problem 3: Stolen Cars*

In many areas of large American cities, street crime is a major problem. Suppose that you live in a neighborhood with a high incidence of automobile thefts. The police estimated that one of every five cars in your neighborhood is stolen each year.

How likely is it that your car will be stolen in the next year? At some time in the next two years? What are your chances of avoiding theft for a three-year period?

How long should a typical resident of your neighborhood expect to wait before her car is stolen? How many years would you estimate it would take for *everyone* in your neighborhood to experience a car theft?

### *2.4 Problem 4: Warning Devices*

Imagine that you are the principal safety officer at a nuclear power plant. You've installed a radiation meter that is supposed to activate a loud warning

alarm if the reactor reaches an unacceptable level of radiation leakage. From your past experience you know that sometimes the alarm rings when the radiation level is far below the dangerous level; you also suspect that under certain circumstances the alarm may fail to go off even when there is a real danger.

How reliable is this safety system? Can we increase the reliability by installing several meters? Suppose three identical meters are put into operation. One morning, one of them sounds an alarm while the other two are silent. Do you close down the reactor? Do you continue operations as usual? What if two of the meters sound warnings and only one is quiet?

### *2.5 A Common Feature of the Four Problems*

Each of these problems involves uncertainty or chance. If 20 percent of the automobiles on your block are stolen each year, then it's possible that everyone will be a victim within five years. But it's also possible that some individuals will never have a car stolen, while others will lose several vehicles to thieves. With good luck, you might never have to report a stolen car; with bad luck, you might become a familiar face at the police precinct house.

If your blood tests positive for the AIDS virus, that might be an instance of a true positive or a false positive. We can't say for sure, but we might be able to estimate a probability or lay odds on the chances that it is a true positive.

We also face the issue of false positives and false negatives in the nuclear reactor example. A false positive means that a warning is issued when none is necessary; a false negative is a situation where the warning device showed no trouble but there was a real danger present. If there are several copies of the warning device in operation, then there are many patterns of successes and failures of the individual devices that could occur, each happening with a different likelihood. Even if every warning device is silent, for example, we can't be absolutely certain that the reactor is operating safely.

In the witchcraft example, we're interested in determining if the historical data justifies a conclusion that gender bias was present. An alternative explanation is that whether or not a given individual was convicted was a matter of chance; the numbers we see may be consistent with a random selection from the pool of men and women.

In order to obtain more precise formulations of these questions so that we can eventually find more exact solutions, we need to develop some of the tools of probability.

By a *probability*, we mean a numerical measure of the chance or relative likelihood that some event will occur. The probability of an event will then be a number between 0 and 1. The closer the probability is to 0, the less likely it is that the event will happen. The closer the probability is to 1, the more likely it is that the event will occur.

An event which has probability 1/2 is just as likely to happen as not to happen. An event with probability 3/4 is three times more likely to happen than an event with probability 1/4.

In the next section we will develop some elementary counting techniques that can help us solve problems in which probabilities play a central role.

# Chapter 3

**Counting**

Our aim in this chapter is to introduce some techniques for counting the number of objects in a particular situation without having to enumerate each one.

## *3.1 Equiprobable Situations and Counting*

At the heart of the solution of many probability problems is a simple calculation involving counting the number of ways certain conditions can be met.

Let's consider first a very simple example. Suppose we take a coin and flip it once. What is the probability that it lands with Heads showing? There's really no satisfactory way of answering this question unless we have some more information about the coin. Is it properly balanced so that there is no bias in the direction of landing Heads or Tails? Or has it been weighted so that one side is more likely to show than the other?

If each side is as likely to show as the other, then we may reasonably talk about our coin being a fair one: For a fair coin, we all have a strong belief that the answer should be one-half. But can we justify it? An easy explanation is that :

— There are only two possible ways the coin can land: Heads or Tails;

— Since the coin is fair, each of the two ways is equally likely;

— In only one of these ways do we have a Head;

— Thus the probability of getting a head is simply the number of ways (1) of getting a Head divided by the number of ways (2) the coin flipping experiment can end: 1/2.

We'll now apply the same type of reasoning to a more complex situation. Take the same fair coin and flip it 4 times, recording the sequence of Heads and Tails. What is the probability of seeing 2 Heads and 2 Tails?

To answer this, we'll use the same four step approach:

Determine how many patterns we might record. In this case there are 16 possible patterns. Using H to represent Heads and T for Tails, we can list the 16:

| | | | |
|---|---|---|---|
| HHHH | HTHH | THHH | TTHH |
| HHHT | HTHT | THHT | TTHT |
| HHTH | HTTH | THTH | TTTH |
| HHTT | HTTT | THTT | TTTT |

— Since the coin is fair, each of the 16 possibilities is as likely to occur as any other outcome.

— Determine how many ways we can obtain 2 Heads and 2 Tails. Of the 16 patterns, only the following six have this feature:

| | | |
|---|---|---|
| HHTT | HTHT | HTTH |
| THHT | THTH | TTHH |

Thus the probability of obtaining 2 Heads and 2 Tails in four tosses of a fair coin is

$$\frac{\text{Number of ways to obtain 2 Heads and 2 Tails}}{\text{Number of ways for experiment to end}} = \frac{6}{16} = \frac{3}{8}$$

Generalizing from these examples, we can formulate an extremely important result in probability: If all possible outcomes of an experiment are equally likely, then the probability of a certain result is the ratio of the number of ways in which that result can occur divided by the number of possible outcomes of the experiment.

Note the important assumption that all possible outcomes must be equally

likely. If a coin has been weighted so that it will land Heads twice as often as Tails, then we could not conclude that there is a probability of 1/2 of getting a Head on a single flip of that coin.

When all the outcomes of an experiment are equally likely, we speak of an *equiprobable* situation. The determination of probabilities in equiprobable situations reduces then to two counting problems:

1. How many different outcomes are there?

2. For how many of these outcomes is the desired condition true?

One possible way to approach these questions in a particular situation is to list every possible outcome of the experiment, count them one at a time, then count one at a time those outcomes in which the desired outcome is true. This is the way we answered our question about the probability of getting 2 Heads and 2 Tails in four flips of a fair coin.

Such an approach is logically possible, but it has severe practical problems. If we want to determine the probability of getting 20 Heads and 20 Tails in 40 flips of a fair coin, then this approach would have us first list all the patterns of 40 Heads and Tails we might observe. There are more than a TRILLION (1,000,000,000,000) such patterns! There must be an easier way to answer such questions!! We explore such a way in the next section.

Before we take up the Fundamental Counting Principle, let's look at one more example of counting by enumeration. Consider a fair pair of dice. Each die is a six-sided cube, labeled with one, two, three, four, five or six spots. By "fair", we mean that each face is as likely to appear on the top when the die comes to rest after being rolled. Thus, each of the numbers between 1 and 6 has probability 1/6 of being rolled. In many games (such as Monopoly and "craps"), a pair of dice is rolled and the outcome depends on the sum of the numbers which appear on the dice. There are 36 different ways the dice could land. We can list them conveniently as pairs of numbers: (a,b) where a is the result from die 1 and b is the result from die 2:

| | | | | | |
|---|---|---|---|---|---|
| (1,1) | (1,2) | (1,3) | (1,4) | (1,5) | (1,6) |
| (2,1) | (2,2) | (2,3) | (2,4) | (2,5) | (2,6) |
| (3,1) | (3,2) | (3,3) | (3,4) | (3,5) | (3,6) |
| (4,1) | (4,2) | (4,3) | (4,4) | (4,5) | (4,6) |
| (5,1) | (5,2) | (5,3) | (5,4) | (5,5) | (5,6) |
| (6,1) | (6,2) | (6,3) | (6,4) | (6,5) | (6,6) |

Since the pair of dice is fair, each of the 36 outcomes is equally likely to occur. Now there are only 11 different possible sums that can be represented,

ranging from 2 to 12. We can list the possible sums and the outcomes which produce them in a simple table:

| Sum | Outcomes | Probability of the sum |
|---|---|---|
| 2 | (1,1) | 1/36 |
| 3 | (1,2) (2,1) | 2/36 |
| 4 | (1,3) (2,2) (3,1) | 3/36 |
| 5 | (1,4) (2,3) (3,2) (4,1) | 4/36 |
| 6 | (1,5) (2,4) (3,3) (4,2) (5,1) | 5/36 |
| 7 | (1,6) (2,5) (3,4) (4,3) (5,2) (6,1) | 6/36 |
| 8 | (2,6) (3,5) (4,4) (5,3) (6,2) | 5/36 |
| 9 | (3,6) (4,5) (5,4) (6,3) | 4/36 |
| 10 | (4,6) (5,5) (6,4) | 3/36 |
| 11 | (5,6) (6,5) | 2/36 |
| 12 | (6,6) | 1/36 |

We see from the table that the most likely sum (the one with the highest probability) is a 7 which has probability 6/36 or 1/6. The probability of rolling a 2,3 or 12 is (1 + 2 + 1)/36 or 5/36. The probability of rolling an odd sum is the sum of the probabilities of rolling a 3 or a 5 or a 7 or a 9 or an 11:

$$\frac{2+4+6+4+2}{36} = \frac{18}{36} = \frac{1}{2}$$

### *3.2 The Fundamental Counting Principle*

Suppose your favorite fast food restaurant features three kinds of sandwiches: chicken, fish and hamburger and a choice of two types of potatoes: French fries and hash browns. How many different types of meals can you order that are made up of a sandwich and some sort of potato?

One way of answering this question is to compile an exhaustive list of all possible menus. In this case, it is easy to list them all:

1. Chicken, French Fries
2. Fish, French Fries
3. Hamburger, French Fries
4. Chicken, Hash Browns
5. Fish, Hash Browns
6. Hamburger, Hash Browns

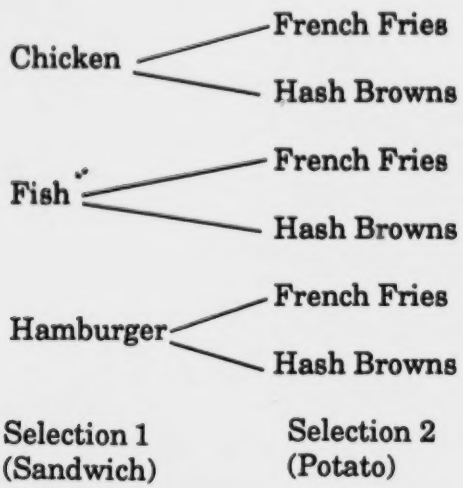
There are six different combinations for this restaurant. But imagine a fast food outlet that boasts 23 varieties of sandwiches and 17 different types of potatoes. Would you want the task of listing all possible menus? This would certainly be an exhausting way to determine how many combinations there are.

Here's another way to think about this problem. Put together your meal by first selecting a sandwich and then a potato. There are three choices for your first selection (the sandwich) and for each of these picks there are two choices for the second selection (the potato). Thus there are (3) (2) = 6 combinations.

Let's try the same reasoning for the expanded fast food operation. Here there are 23 ways of making the first selection and for each of these picks there are 17 choices for the second selection. This gives (23) (17) = 391 combinations.

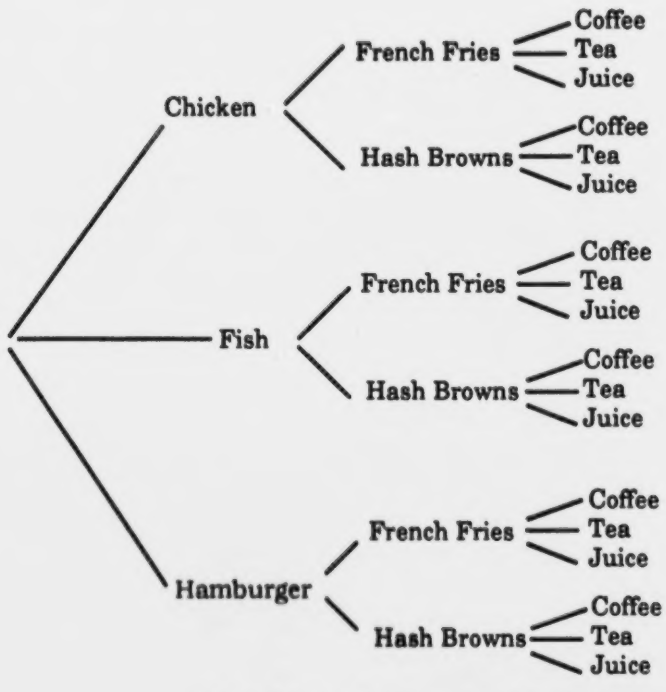
There's a Counting Principle that covers both these cases. It says: if there are M ways of taking a first selection and for each of the picks there are N choices for the second selection, then there are MN combinations of the two selections.

We can draw a helpful picture to illustrate this counting principle:




This kind of a figure, which begins at a root and branches out, is called a *tree diagram*. Many people find tree diagrams a very useful tool in analyzing many different kinds of problems. They are especially helpful in building and analyzing probability models.

To expand on our fast food example, let's complete your meal by selecting a beverage. This restaurant offers three beverages: coffee, tea or juice. We can incorporate this additional level of choice by extending our tree diagram:




For each of our 6 combinations, there are three ways to make the beverage selection. This gives us (6) (3) = 18 different combinations overall.

A General Counting Principle: If one thing can be done in exactly R different ways, and for each of these a second thing can be done in exactly S different ways, and for each of the first two, a third thing can be done in exactly T different ways, and so on, then the sequence of things can be done in exactly RST... different ways.

As another simple example of the use of the General Counting Principle, recall the example of flipping a coin 4 times. The result of each flip is one of two possibilities: a Head or a Tail. For each of the 2 different outcomes of the first flip, there are 2 different outcomes for the second, and in each of these 2 there are still 2 outcomes for the third flip and for both of these there are 2 possibilities for the fourth flip. By the General Counting Principle, there are altogether:

(2) (2) (2) (2) = $2^4$ = 16 possible outcomes

If we flip the coin 40 times, there would be (2) (2) (2) ... (2) = $2^{40}$ possible outcomes

To obtain a rough estimate of the size of $2^{40}$, note first that $2^{10} = 1024$ which is approximately 1000 or $10^3$.

Then $2^{40} = 2^{10}\ 2^{10}\ 2^{10}\ 2^{10}$

which is roughly $10^3\ 10^3\ 10^3\ 10^3$

or $10^{12}$; that is, a 1 followed by 12 zeros.

# Chapter 4

**Permutations and Combinations**

In this chapter, we extend the use of the Fundamental Counting Principle to situations dealing with rearranging or combining objects into collections.

## *4.1 Arranging Objects in a Row*

As another illustration of the use of the Counting Principle, let's determine the number of different ways that Peter, Bob, John, Claudia and Priscilla can arrange themselves in a line for a group photograph for the newspaper. The paper will print the names of the people in the photo in the standard left to right order:

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|

We have 5 distinct ways of making the choice for our first selection, the person who will be on the far left of the photo.

For each of these choices, any of the remaining 4 persons can be picked. After the first two selections, we are free to chose any of the remaining 3 people for the third selection. Hence the number of distinct ways of making

the first three selections is 5*4*3 = 60. [We're using the symbol * to represent multiplication; thus 5*4 is shorthand for the expression you may have seen in the form (5)(4) or 5 x 4.]

To complete our picture, we must make a fourth selection and then a fifth selection. For each of the 60 ways of picking the first three, there are 2 ways of making the fourth selection. For the final selection, only 1 person is left so we have 1 choice. Thus the number of different ways of arranging the 5 people is 5*4*3*2*1 = 120.

Notice that this number is the product of the whole numbers from 1 to 5. Mathematicians have invented the notation 5! as shorthand for this product. If N is any positive whole number, then N! is an abbreviation for the product of the whole numbers from 1 to N. The term N! is read "N factorial." The number N! counts the distinct ways in which N distinguishable objects can be arranged in a row. As some simple examples,

3! = 3 * 2 * 1 = 6
7! = 7 * 6 * 5 * 4 * 3 * 2 * 1 = 5040
1! = 1
8! = 8 * 7! = 40320

For technical reasons we'll explain in a few minutes, 0! is defined to be equal to 1.

### *4.2 Selecting a Group From a Larger Collection*

Let's ask a related, but just slightly harder question. To our original quintet of Peter, Bob, John, Claudia and Priscilla, we add Bruce and David so our group now numbers seven. How many different photographs can we make of a quartet chosen from the seven?

Again, the General Counting Principle gives us the answer. There are 7 choices for the first person, and for each of these there are exactly 6 choices for the second person, then for each of these exactly 5 choices for the third and, finally, for each of these, there are 4 choices for the fourth person. The number of distinct photographs is

7 * 6 * 5 * 4 = 840

Observe that the left hand side of this equation looks like the "high end" of 7!. In fact, we can write 7*6*5*4 as:

$$7 * 6 * 5 * 4 = \frac{7 * 6 * 5 * 4 * 3 * 2 * 1}{3 * 2 * 1} \qquad \frac{7!}{3!}$$

In one of these photographs, we see as we go from left to right Peter, Priscilla, Claudia and Bob. In a different photo, these same four appear arranged in a different order: Bob, Priscilla, Peter and Claudia. In how many distinct photos does the same group appear? The answer is easy: the number of ways that 4 people can be arranged in a line; that is, 4! or 24. If we were to consider another group of four (say Bob, Claudia, David and Bruce), there would also be 24 distinct photos containing them.

We see then that the 840 photographs can be broken up into piles of 24, with a separate pile for each quartet. How many quartets would there be? Simply divide the total number of photos by the number of photos in each pile:

840 / 24 = 35

It's useful to rewrite this equation in a way which recalls something of its history:

$$\text{Number of different groups of 4 that can be chosen from a set of 7} = \frac{7 * 6 * 5 * 4}{4 * 3 * 2 * 1} = \frac{7!}{4!\,3!}$$

If you trace through the reasoning behind this example applied to the problem of selecting 6 different people from an original group of 10 to form an executive committee, you will find that

$$\text{Number of distinct ways to choose 6 objects from a set of 10 objects} = \frac{10!}{6!\,4!} = 210$$

On the other hand, if we wish to count the number of ways we could fill six particular offices — President, Vice-President, Recording Secretary, Corresponding Secretary, Treasurer, and Auditor — then there would be many more ways, namely $10 * 9 * 8 * 7 * 6 * 5 = 151{,}200$

In this example, we are not simply choosing six people to form a group, but rather selecting six to fill particular, distinguishable positions. Probability theorists often speak of selection considering order or ignoring order. In

selecting an Executive Committee of 6 persons, we are ignoring order; it doesn't matter which person is chosen first, which one second, etc. All those chosen serve in the same capacity. But in filling specific offices, we are considering order.

If the first person selected will be President and the second one Vice-President, then the role you play in the organization very much depends on when you were chosen. Whenever we talk about the number of ways of choosing a certain size group of objects from a larger collection, we will mean that we are ignoring order.

### *4.3 N Choose K*

In general the number of distinct ways to choose K objects from a set of N distinguishable objects is

$$\text{Number of distinct ways to choose K objects from a set of N objects} \qquad \frac{N!}{K!\ (N-K)!}$$

There are so many situations where it is useful to keep track of such a number that mathematicians have invented a special symbol for this quotient:

$$\binom{N}{K} = \frac{N!}{K!\,(N-K)!}$$

This symbol is read "N choose K" since it is the number of different ways we can choose K objects from a collection of N objects.

One way to convince yourself of the validity of this formula is to consider that in choosing K objects from a set of N, we are simultaneously picking N-K objects not to be in our collection. As an example, in choosing 6 executive committee members from an organization of 10 members, we can either designate the specific 6 committee members or designate 4 organization members who will not be on the committee. Including K members on a committee is equivalent to excluding N-K members. Thus the number of distinct ways to choose K objects from a set of N should be equal to the number of ways of choosing N-K objects from a set of N. Our formula should produce the same result. Let's check this. By the formula, the number of

ways of choosing N-K objects from a set of N is

$$\frac{N!}{(N-K)!\ (N-(N-K))!} = \frac{N!}{(N-K)!\ K!}$$

so that "N choose K" and "N choose (N-K) " are — as they should be — equal.

It's always useful to check a formula in extreme cases. Does this formula give the correct answer, for example, for the number of ways of choosing 1 object from a set of 10? Of course, we don't need any formula to answer this question. Since any of the 10 can be selected as the unique object, there are 10 distinct ways. The formula gives

$$\binom{10}{1} = \frac{10!}{1!\ 9!} = \frac{10 * 9!}{1 * 9!} = 10$$

What about an even more extreme case? How many ways can we select 10 objects from a collection of 10? Obviously, only 1 way since all 10 must be included. The formula gives

$$\binom{10}{10} = \frac{10!}{10!\ 0!} = \frac{1}{0!}$$

The only way this formula gives the correct answer is if we define, as we did do earlier, 0! to be equal to 1.

To gain a little more insight into this result, suppose we want to count the number of ways to select 5 letters from the collection A,B,C,D,E,F,G,H of 8 letters. Our formula says the answer is "8 choose 5". Another way to look at the problem is that the task of selecting 5 letters to keep is equivalent to the job of picking 3 letters to discard: every choice of 3 to discard defines a collection of 5 to keep. Thus the number of ways to select 5 objects from a collection of 8 should be equal to the number of ways to select 3 objects from 8. Since this latter number is "8 choose 3", we should have

8 choose 5 = 8 choose 3

OR

$$\binom{8}{5} = \binom{8}{3}$$

$$\frac{8!}{5!\,3!} = \frac{8!}{3!\,5!}$$

which we see is true.

Here's another example:

$$\binom{20}{7} = \frac{20!}{7!\,13!} = \frac{20*19*18*17*16*15*14*13!}{7*6*5*4*3*2*1*13!}$$

$$= \frac{(5)(4)(19)(6)(3)(17)(2)(2)(2)(2)(3)(5)(2)(7)}{(7)(6)(5)(4)(3)(2)(1)}$$

$$= \frac{(5)(4)(19)(6)(3)(17)(2)(2)(2)(2)(3)(5)(2)(7)}{(7)(6)(5)(4)(3)(2)(1)}$$

(* & % ! ^ # mark factors in the numerator and denominator that cancel each other)

$$= 19 * 17 * 2 * 2 * 2 * 3 * 5 * 2$$

(we used special symbols to show factors in the numerator and denominator of the fraction which cancel each other)

The number represented by

$$\binom{N}{K} = \frac{N!}{K!\,(N-K)!}$$

can be calculated by hand for small values of N and K. For larger values it is quicker and accurate to use a computer. Appendix A contains a short BA-

SIC program to evaluate "N choose K."

With the help of this program, we evaluated 40 choose 20 to be about $1.378465*10^{11}$. Thus the probability of obtaining exactly 20 Heads (and therefore 20 Tails) on 40 flips of a fair coin is

$$\frac{\binom{40}{20}}{2^{40}}$$ which is approximately equal to .125

# Chapter 5

**Probability Notation and Basic Rules**

## *5.1 Notation*

Let A represent an outcome or set of outcomes for some probabilistic situation. For example, if we toss a fair coin twice, A might represent the elementary outcome "Head on the first flip, Tail on the second" or the more complex outcome "At least one Head." The latter outcome may be broken up into the elementary outcomes "Head on First, Head on Second", "Head on First, Tail on Second", "Tail on First, Head on Second."

We will let A* represent the reverse or complement of A. This is often read "not A" and is the outcome consisting of all situations where A is not true. For example, if A is the outcome "At least one Head", then A* is the outcome "No Heads."

We'll use Pr(A) to represent the probability that A occurs. Then Pr(A*) is the probability that A does not occur; alternatively, you can think of Pr(A*) as the probability that "not A" occurs.

In many situations, we'll want to deal with the probability that two events, say A and B, both occur. This is usually denoted by Pr(A and B) but some people replace the word "and" by a comma to obtain Pr (A,B). If A represents passing anthropology and B stands for passing biology, then Pr(A,B) is the probability of passing both anthropology and biology. In this example, what would Pr(A,B*) mean? The probability of passing anthropology and not passing biology.

If D and E are two events, then the notation Pr(E/D) means the *conditional probability of E given D.* This measures the relative likelihood of E occurring given that you know that D is true.

For example, if our experiment consists of rolling a fair die, we might let E represent the outcome "a 6 is rolled" and D the outcome "an even number is rolled." Then the probability of E occuring, Pr(E), is just 1/6. But if we are told that an even number has been rolled and are then asked how likely it is that it was a 6, our probability estimate would change. We know that one of three outcomes (2,4,6) has happened, each of which is equally likely, so our proper response is 1/3. Thus Pr(E/D) = 1/3.

Note that in this example, Pr(D) is 1/2 so that

$$\Pr(E/D) = \frac{\Pr(E,D)}{\Pr(D)} = \frac{1/6}{1/2} = 1/3$$

It can be shown, more generally, that if D is any event with positive probability, then

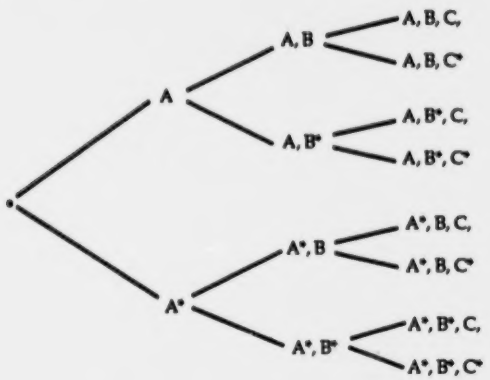
$$\Pr(E/D) = \Pr(E,D) / \Pr(D)$$

or, equivalently,

$$\Pr(E,D) = \Pr(E/D)\ \Pr(D)$$

## *5.2 Basic Rules of Probability*

There are four essential rules for dealing with probabilities. These can be easily derived and justified by considering a simple tree diagram.




In our tree diagram, we often label a branch with the conditional probability associated with the transition from the left hand end to the right hand end.

1. The Probabilities-Must-Add-To-One Rule, Part I: The probabilities for all the branches coming out from a single point in the tree must add up to 1. For example,

$Pr(A) + Pr(A^*) = 1$ and

$Pr(B \mid A) + Pr(B^* \mid A) = 1$

2. The Probabilities-Must-Add-To-One Rule, Part II: The probabilities for all the terminal outcomes on a probability tree must add up to 1. In our diagram,

$Pr(A,B,C) + Pr(A,B,C^*) + Pr(A,B^*,C) + Pr(A,B^*,C^*) +$
$Pr(A^*,B,C) + Pr(A^*,B,C^*) + Pr(A^*,B^*,C) + Pr(A^*,B^*,C^*) = 1$

3. The Multiplication Rule: To find the probability for any terminal outcome, multiply together the probabilities for the different outcome branches along the path from the initial root to that terminal outcome.

More generally, to find the probability of reaching any particular point in the tree, find the path from the root to that point and multiply together the probabilities for the branches that make up that path. For example,

$Pr(A,B ,C^*) = Pr(A) * Pr(B \mid A) * Pr(C^* \mid B)$
$Pr(A,B^*) = Pr(A) * Pr(B^* \mid A)$
$Pr(A^*,B^*,C) = Pr(A^*) * Pr(B^* \mid A^*) * Pr(C \mid B^*)$

4. The Addition Rule: To find the probability of an outcome corresponding to more than one terminal point, add up the probabilities of all these terminal points.

More generally, to find the probability of reaching any particular point in the tree, add up the probabilities for all paths from the root to that point. For example,

$Pr(B,C^*) = Pr(A,B,C^*) + Pr(A^*,B,C^*)$
$Pr(C) = Pr(A,B,C) + Pr(A,B^*,C) + Pr(A^*,B,C) + Pr(A^*,B^*C)$
$Pr(B) = Pr(A,B) + Pr(A^*,B)$

# Chapter 6

**Expected Value**

## *6.1 Assessing a Gamble*

A "friend" offers you the following gamble. He will provide a coin which you get to flip. If the coin lands Heads, he will pay you $3. If the coin lands Tails, you must pay him $2. If the coin is fair, that is, equally likely to land Heads as Tails then you would enthusiastically participate in the gamble. You would win on about as many flips as you would lose; the payoff for winning is greater than the penalty for losing. The gamble is stacked in your favor. Can we measure how favorable it is to you?

Would the gamble still be favorable if the coin has been weighted so that Heads appear only with probability .45? What if Heads only come up 40 percent of the time?

One natural way to respond to such questions is to consider what our average profit per flip would be if we spent the evening gambling in this pleasant manner. We'll start our analysis with the fair coin. Suppose we flip it 100 times. Then we would expect to see Heads about 50 times and Tails about 50 times. Thus we would receive a total of about (50) ($3) = $150 and pay out a total of about (50)($2) = $100. Our total profit would be about $150 - $100 = $50. This would amount to an average profit of about $50/100 or 50 cents per flip.

If we flip the coin 200 times, then our estimated net profit would be (100)($3) - (100)($2) = $100, but the average profit per flip would remain the same: 50 cents.

If we flip the coin N times, then our estimated net profit would be (N/2) ($3) - (N/2) ($2) = $N/2. The average profit per flip would be ($N/2) / N which is still 50 cents. We can also write this calculation as

$$\text{Average Profit Per Flip} = \frac{(N/2)(\$3) - (N/2)(\$2)}{N} = \frac{1}{2}(\$3) + \frac{1}{2}(-\$2)$$

= ($3) (Probability of Head) +
(-$2) (Probability of Tail)

The average profit per flip is called the *expected value* of the gamble. It is a measure of what you would expect as the average profit of the gamble to you if you were to repeat the gamble a large number of times. For this example, the expected value is 50 cents. Note that on any particular flip of the coin, the result is either a gain of $3 or a loss of $2. There is no single flip on which you gain 50 cents. But if you were to spend the evening playing this game, you would expect over the "long haul" to average half a dollar gain per flip.

### *6.2 Expected Value for Two Outcome Experiment*

If an experiment or gamble has two possible numerical outcomes or payoffs a, which occurs with probability p,and b, which occurs with probability l-p, then the *expected value* is defined to be

$$EV \quad ap + b(1-p)$$

In our example, a = 3, b = -2 and p = 1/2, so that EV = 1/2. If the coin is weighted so that p, the probability of a Head, is .45, then

$$EV = 3(.45) + (-2)(.55) = 1.35 - 1.10 = .25$$

For such a gamble, the expected value per flip is 25 cents. The gamble is still favorable to you but not as advantageous as it would be with a fair coin.

If the coin is so unbalanced that Heads appear only 40 percent of the time, then p = .4 and the expected value is

$$EV = 3(.4) + (-2)(.6) = 1.2 - 1.2 = 0$$

The expected value per flip here is 0. You should expect, in the long run, to "break even" on such a gamble. We would call such a gamble a *fair* bet.

If the probability p of a Head drops below .4, then the expected value will become negative. You would expect a net loss on the average per flip.

### *6.3 Expected Value for Multiple Outcome Experiment*

If an experiment has many different possible numerical outcomes, then the *expected value* is defined as the sum of the products of each numerical outcome multiplied by the probability that the particular outcome occurs.

For example, suppose your aunt is the director of the personnel office at a large university. She agrees to hire you for a summer job, but isn't sure now exactly what position she can give you. She gives you the following estimates:

| Position | Summer Salary | Probability of Getting Job |
|---|---|---|
| Custodian | $1200 | .4 |
| Dishwasher | $1350 | .3 |
| Painter | $1425 | .2 |
| Computer Programmer | $2200 | .1 |

Then your expected summer salary would be

($1200)(.4) + ($1350)(.3) + ($1425)(.2) + ($2200)(.1) =
$480 + $405 + $285 + $220 = $1390

If we flip a fair coin three times and record the number of Heads, then there are four possibilities:

0 Heads if outcome is TTT
1 Head if outcome is HTT, THT, or TTH
2 Heads if outcome is HHT, HTH, or THH
3 Heads if outcome is HHH

In this case, the expected number of Heads is

0 (1/8) + 1 (3/8) + 2 (3/8) + 3 (1/8) = 12/8 = 1.5

As a third example, we'll calculate the expected value of the sum when rolling a pair of fair dice. Here we have an experiment with 11 different numerical outcomes: 2,3,4,...,12. Recalling our table of probabilities in an earlier section, we have:

Expected Value of Sum = (2) (Probability sum is 2)
+ (3) (Probability sum is 3)
+ ...
+ (12)(Probability sum is 12)

(2) (1/36) + (3) (2/36) + (4) (3/36) + (5) (4/36) +
(6)(5/36) + (7)(6/36) + (8)(5/36) + (9)(4/36)
+ (10) (3/36) + (11) (2/36) + (12) (1/36)

= (2+6+12+20+30+42+40+36+30+22+12)/36 = 58/9 = 6.4

### *6.4 Expected Value For Infinitely Many Outcome Experiment*

How many times should we expect to have to roll a die before the number 4 appears?

The probability of getting a 4 on a single roll is 1/6. We might get a 4 on the first roll. This happens with probability 1/6. The next possibility is that we don't get a 4 until the second roll. For this to happen, we must obtain a "non-four" on the first roll (that is, 1,2,3,5 or 6) and a 4 on the second roll. The probability of a "non-four" is 5/6, so that the sequence

Non-Four, Four

has probability

(5/6) (1/6)

Next, we must consider the possibility that the first Four appears on the third roll. This happens when the sequence is

Non-Four, Non-Four, Four

which occurs with probability

(5/6) (5/6) (1/6) = $(5/6)^2$ (1/6)

You should see a pattern beginning to emerge. For the initial Four to appear on the kth roll, we must have a sequence of (k-l) Non-Fours followed by a Four:

Non-Four, Non-Four, ..., Non-Four, Four

which occurs with probability

$(5/6)\ (5/6)\ \ldots\ (5/6)\ (1/6) = (5/6)^{(k-1)}\ (1/6)$

Now we're set to write down an expression for the expected value V of the number of rolls before a Four appears:

V = (1) (Probability that Four first appears on Roll 1) +
(2) (Probability that Four first appears on Roll 2) +
(3) (Probability that Four first appears on Roll 3) +
... +
(k) (Probability that Four first appears on Roll k) +
...

$V = (1)(1/6) + 2(5/6)(1/6) + 3(5/6)^2\ (1/6) + 4(5/6)^3\ (1/6) + \ldots$

$V = (1/6)\ [1 + 2(5/6) + 3(5/6)^2 + 4(5/6)^3 + \ldots]$

$V = (1/6)\ B$

where B is the value of the expression inside the square brackets. Replace 5/6 with x inside the brackets to obtain:

$B = 1 + 2x + 3x^2 + 4x^3 + \ldots$

Now we can appeal to a mathematical formula (which we derive in Appendix II) that if x is any number strictly between -1 and 1, then the number on the right hand side is $1/(1-x)^2$

and when x = 5/6, this becomes

$$1/(1 - (5/6)\ )^2 = 1/(1\ /6)^2$$

so we see that

$V = (1/6)\ B = (1/6) / (1/6)^2 = 1/(1/6) = 6$

Although the derivation is a bit long, the result is simple to state and understand: If a particular outcome of an experiment has probability 1/6, then the expected number of times the experiment must be repeated before that outcome occurs has value 6.

There is no special magic about "1/6" here. If we go through the calculations above with an arbitrary probability p replacing 1/6, then the value of V works out to 1/p. We may state this important result as:

> If a particular outcome of an experiment has probability p, then the expected number of times the experiment must be repeated before that outcome occurs has value 1/p.

As another example of this result, suppose you have a key chain with 10 keys on it, only one of which will open the front door of your house. It is late at night when you return home. There is no appreciable moon light and the street lamps on your block are off due to a power failure which has also cut off electricity to your house. In other words, it's too dark for you to see clearly which key is which. You grab a key at random and try it in the door. There is a probability of 1/10 that you picked the right key.

Thus, if this situation occurs repeatedly, you will select the correct key on the first try roughly one in every ten tries. Of course, if you don't succeed on the first selection, you'll want to try another key. But suppose your hand is not very steady and you drop the key chain on the ground after each unsuccessful try. Then we can expect on the average to make 10 tries before hitting on the correct key.

Note that in this situation, it is possible to try the wrong key several different times in you quest for the right key. This is an example of sampling with replacement: you select one object from a larger population, record the one selected, return it to the larger set and then select again.

### *6.5 Expected Value When Sampling Without Replacement*

In other situations, you may be faced with sampling without replacement: you select one object from a larger population, record its value and discard it, and then select another object from the larger population. In sampling without replacement, there is no possibility of selecting the same object twice; the size of the larger population decreases by one after each selection. In sampling with replacement, the size of the larger population remains fixed during the entire selection process; we might choose some objects several times.

As an example of sampling without replacement, suppose that you're trying to guess the correct answer to a multiple choice question. Four suggested answers are given. Only one of them is correct. You may make repeated guesses. After each guess, you're told if you are right or wrong. How many guesses would you expect to make?

Let's assume that when you guess the correct answer the game will end. We'll also assume that you won't repeat a guess that you've been told is a wrong one. This means that you will certainly guess the correct answer by the fourth try. You might get it as early as the first try. The expected number of guesses is:

(1) (Probability you guess on first try) +
(2) (Probability you guess on second try) +
(3) (Probability you guess on third try) +
(4) (Probability you guess on fourth try)

The probability that you guess correctly on the first try is simply 1/4 since there are four equally likely choices, but only one correct one.

In order for you to guess correctly on the second try, your first guess must be wrong but your second guess must be right. The probability that the first guess is wrong is 3/4. After making the first guess, you're told that it is incorrect. That means that one of the remaining three answers must be the right one. So on the second try, you are guessing at random from a collection of three objects; the probability of a success is 1/3. Thus the probability of guessing correctly on the second try is

Probability (Incorrect, Correct) = (3/4)(1/3) = 1/4

In order for you to guess correctly on the third try, your first two guesses must be wrong, but your third guess must be right. We're looking for

Probability (Incorrect, Incorrect, Correct) = (3/4) (2/3) (1/2) = 1/4

The last factor (1/2) comes from the consideration that your first two incorrect guesses has narrowed the set of possibly correct answers down to 4 - 2 = 2 and that you will pick one of these two at random.

Finally, the only way not to guess correctly until the fourth try is to have three incorrect guesses followed by the right answer:

Probability (Incorrect, Incorrect, Incorrect, Correct)

= (3/4) (2/3) (1/2) (1) = 1/4.

Thus, the expected number of guesses is

1(1/4) + 2 (1/4) + 3 (1/4) = 4 (1/4) = 10/4 = 2.5

You may be interested in working out the expected number of keys you

must try to open your front door if we never try a second time any key which didn't work the first time. The reasoning is similar to guessing the right answer to a multiple choice question. The expected number is 5.5 keys.

# Chapter 7

**The Witchcraft Problem**

In 17th century America, a good many people were accused of witchcraft. Some received formal trials; most of these were found not guilty. Both men and women faced witchcraft charges, but not in equal numbers. Historians wish to know if there was a sex bias in these trials: were the women more likely to be convicted than the men?

## *7.1 Some Numerical Evidence*

For a fascinating study of witchcraft, see *Entertaining Satan: Witchcraft and the Culture of Early New England* by John Putnam Demos, publixhed by Oxford University Press in 1982. In the table below there is some data Professor Demos collected about witchcraft trials in New England between 1630 and 1690. Of the 89 cases brought to trial, there were 68 accused women and 21 accused men. Fourteen of the women and two of the men were convicted. We arrange the data in a table with variables GENDER and CONVICT:

| | CONVICT | | |
|---|---|---|---|
| GENDER | NO | YES | TOTAL |
| MALE | 19 | 2 | 21 |
| FEMALE | 54 | 14 | 68 |
| TOTAL | 73 | 16 | 89 |

Examine this data for a moment. Were women more likely than men to be *accused* of witchcraft? Were women more likely than men to be *convicted*? Are the answers "obvious" from the data?

One way to get a better understanding of the witchcraft problem is to convert the absolute frequencies into relative frequencies or percentages. By the absolute frequency of an event, we mean the *number* of times it occurred. By the relative frequency we mean the *proportion* of times this event happened among all observations. Of the 21 men charged with witchcraft, only 2 were

convicted. Thus the absolute frequency of convictions is 2. This gives a relative frequency or percentage of (2/21) * 100 or 9.5 percent. Similarly, the percentage of women convicted (54 of 68) is 54/68 * 100 or 79.4 percent. The table below shows the percentages for all categories:

| | CONVICT | | |
|---|---|---|---|
| GENDER | NO | YES | TOTAL |
| MALE | 90.5 | 9.5 | 100 % |
| FEMALE | 79.4 | 20.6 | 100 % |
| TOTAL | 82.0 | 18.0 | 100 % |

Does the presentation of the data in this format make it seem more likely that there was bias against women? Or does it look less likely that there was sex discrimination?

Perhaps you focused on the 'YES' column. Here we observe that the rate of conviction for women (20.6%) is more than twice as great as the conviction rate (9.5%) for men. It does appear from this view that women suffered a disproportionately high rate of conviction.

But if we examine the 'NO' column, we may reach a different conclusion. Here we note that the percentage of males found innocent (90.5%) is only 11 percentage points higher than the figure(79.4%) for women. Moreover, the overall acquittal rate for the entire accused group is 82 percent, just slightly more than the rate for the women. Perhaps the data indicates that men and women were treated equally?

Is the difference between the conviction rate for men and that for women large enough to be significant? After all, there times as many women were accused as men. Perhaps this pattern of conviction is about what we should expect to see. On the other hand, seven times as many women as men were convicted. Maybe this pattern did not occur by chance? How can we resolve these questions? Is there a simple way to determine whether or not a real relationship between gender and the outcome of the accusations?

### *7.2 What Patterns Occur By Chance?*

As a first approach to this problem, let's see what patterns might occur by chance. Then we can compare these to the historical data. We want to test an hypothesis: women were more likely than man to be convicted of witch-

craft. If patterns as extreme or more than the historical data are relatively rare, then we may conclude that our hypothesis is correct. If such patterns are reasonably common, then we have no grounds for accepting the hypothesis as an explanation for the observed historical data.

What do we mean by the chance process? There are a total of 89 people in our sample population of accused witches. Of this number, 21 were Male and 68 were female. We also know that 16 were convicted. If there was no gender bias, then each of the 89 people had the same chance to be selected as "guilty" party. By a chance process here, we mean one that will select 16 of the 89 people at random so that each accused person has the same probability of selection as every other accused person. We will note the sex of each person selected and keep track of the total number of men and women chosen.

### *7.3 Simulating a Chance Process*

There are many ways to simulate this chance process. We could take 89 index cards. Label 21 with the word "Male" and 68 with the word "Female." Shuffle the cards together and then deal out 16. Alternatively, you could put 21 green jelly beans and 68 red ones into a bag, shake the bag several times and then reach in and select—without looking—16 of the candies. If you don't have access to a computer, we suggest that you try one of these experiments. [The jelly bean simulation can end with eating the experimental apparatus!]

A computer simulation of this chance process is easy to create. The computer simulation has many advantages. It is easy to repeat the experiment many hundreds—or even thousands of times—in a few moments. So it is much faster (and less fattening) than using jelly beans.

The output of a typical run of this program is:

There were 3 men and 13 women selected.

There's not very much we can conclude from a single run of the program. To get a better sense of what typical patterns of choices result from this simulation, we need to run the program a larger number of times. In the table below, we show the results of 2 different sets of repetitions of the program. Each set shows the frequencies that were seen in 1000 runs of the program.

Results of 1000 repetitions of WITCHES program

| Number of Women Chosen | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Frequency (Sample 1) | 0 | 3 | 11 | 47 | 90 | 189 | 206 | 219 | 152 | 70 | 13 |
| Frequency (Sample 2) | 1 | 3 | 15 | 31 | 108 | 159 | 244 | 225 | 140 | 62 | 12 |

In the total of 2000 simulations of choosing 16 people from the group of 89, we never wound up with fewer than 6 women. We did obtain a total of six instances in the 2000 where exactly 7 women and 9 men were picked. It is possible then for purely chance drawings with no gender bias to result in cases where more men than women are selected. The simulation leads us to believe, however, that such cases are extremely rare. If the historical pool of 89 accused witches did contain only 21 men and if the convicted 16 had a majority of male members, we might reasonably conclude that there was gender bias in the conviction process. But what about the historical data we have that showed 14 women and 2 men in the convicted group?

Note that in the first sample of 1000 simulations of the 16 draws, we had 235 (152 + 70 + 13), cases when 14 or more women were selected. In the second sample, we saw 214 such cases.

From our simulations, it would be reasonable to conclude that if individual men and women had equal chances of being selected to be one of the 16 "guilty" parties, we would expect to see in slightly more than 20 percent of the cases situations when 14, 15 or 16 were women.

Since results as extreme, or more extreme, than the ones recorded in the historical data would appear by chance roughly one in five times—a not very rare proportion—there is no sufficient evidence that gender bias existed.

### *7.4 Using Probabilistic Reasoning*

Can we obtain a more precise estimate of the probability that 14 or more women would be chosen? Yes. Probability theory gives us the tool to calculate this number exactly. The outcome of obtaining at least 14 women in 16 can be broken up into three mutually exclusive outcomes: exactly 14 women, exactly 15 women, and exactly 16 women. Thus

$$\begin{aligned}\Pr(\text{14 or more women}) = {} & \Pr(\text{exactly 14 women}) + \\ & \Pr(\text{exactly 15 women}) + \\ & \Pr(\text{exactly 16 women})\end{aligned}$$

objects, without replacement, from a collection of 89 objects each of which has the same chance of being picked. We know that the sampling is without replacement since 16 different objects must be selected. Thus

$$\Pr(\text{exactly 14 women}) = \frac{\text{Number of ways 14 women and 2 men can be selected}}{\text{Number of ways 16 people can be selected from 89}}$$

$$= \frac{\binom{68}{14}\binom{21}{2}}{\binom{89}{16}}$$

The probability of exactly 15 women being selected is a similar fraction. The denominator is the same, but the numerator would be:

$$\binom{68}{15}\binom{21}{1}$$

Can you see now what is the probability that all 16 chosen people are women? The probability that 14 or more women will be picked is the sum

$$\frac{\binom{68}{14}\binom{21}{2} + \binom{68}{15}\binom{21}{1} + \binom{68}{16}\binom{21}{0}}{\binom{89}{16}}$$

Although we could, in principle, find the value of this fraction with pencil and paper calculations, the necessary computations for evaluating some of these binomial coefficients are so arduous that we turn to the computer for help. The computer calculates the value of this fraction to be .2077 so we see again that the probability that at least 14 women will be chosen—if gender plays no role in the selection—is slightly more than 20 percent. Note that our simulation results are quite close to this value.

# Chapter 8

**Screening For AIDS**

## *8.1 The AIDS Disease*

Acquired immune deficiency syndrome (AIDS) is a terrifying contagious disease which many observers believe has already reached epidemic proportions in the United States. In October, 1986 the National Academy of Sciences said that the AIDS problem required "perhaps the most wide-ranging and intensive efforts ever made against an infectious disease....If the spread of the virus is not checked, the present epidemic could become a catastrophe" (*The New York Times*, October 30, 1986))

AIDS is invariably a fatal illness; no diagnosed AIDS victim has lived more than 5 years. By October, 1986 more than 12,000 Americans had died from AIDS. Medical experts predict that AIDS will claim 100,000 more lives by the end of this decade. Dr. Ward Cates of the federal Center for Disease Control has warned

> Anyone who has the least ability to look into the future can already see the potential for this disease being much worse than anything mankind has seen before. (James I. Slaff and John K. Brubaker, *The Aids Epidemic*, New York: Warner Books, 1985, Page 4)

AIDS is transmitted from one person to another primarily through sexual relations, intravenous drug use, and blood transfusions. To halt the disease's spread, we need first to identify the carriers of the AIDS virus so that they may take proper precautions. Scientists have been searching for an effective method to identify AIDS virus carriers. Such methods have been found for other diseases; for example, chest x-rays are routinely given now to detect tuberculosis.

## *8.2 Blood Tests*

One promising technique is a blood test (ELISA) which detects the presence of certain antibodies that the body produces in reaction to the AIDS virus. While such a test may be extremely accurate, it is subject to two possible types of errors. First, it may fail to detect the illness in someone who actually suffers from it. Secondly, it may "detect" AIDS in a person who really doesn't have the disease.

Suppose the blood test is refined to the point where it correctly identifies AIDS in 95 percent of the population which has the disease; thus 5 percent of those who have AIDS get a "false negative" result from the test. Further-

more, suppose that 99 percent of those tested who are free of the AIDS virus also receive a negative test result. This means that 1 percent of AIDS-free individuals will receive a "false positive" test result.

Many people have argued that the government should fund a mass screening program to detect AIDS carriers. Suppose such a program is undertaken and all individuals have this blood test. The actual number of AIDS virus carriers in the United States is unknown, but it is estimated that approximately 1 person per thousand is so afflicted. If your blood is tested and the result is positive, how worried should you be? What are the chances that you actually have AIDS?

## *8.3 An Informal Solution*

We're assuming that the ELISA test (Enzyme-Linked ImmmunoSorbent Assay) correctly points to the presence of AIDS virus in 95 percent of those who actually have the virus and that it incorrectly identifies the presence of the virus in only 1 percent of those who really don't have it. We've also assumed that approximately one in every thousand persons in the population actually carries the AIDS virus.

Suppose then we give the ELISA blood-screening test to 100,000 randomly chosen people. We would expect that about 100 of these people have AIDS. About 95 of these 100 will receive positive test results. Of the 99,900 people who do not have AIDS, we expect that 1 percent of them, or 999, will receive a false positive. There are 95 + 999 = 1094 persons with positive tests, of which only 95 truly have the AIDS virus. Thus, among those who have positive tests, the probability of having AIDS is

95/1094 = .087

which is less than 9 percent. We would conclude that more than 90 percent of those who receive a positive test that indicates they have the AIDS virus are actually free of the virus! The vast majority of people who become terrified when told they have a positive test for AIDS are needlessly frightened.

What can we do to prevent this unnecessary panic? There are several approaches we might take.

## *8.4 Improving the Procedure*

### 8.4.1 Sensitivity

First, try to improve the accuracy of the test. For an illness like AIDS which can be spread easily, we definitely do not want our screening procedure

to result in many false negatives. By the sensitivity of a screening test or procedure, we mean the probability that a person who has the disease has a positive test result. So we might concentrate our efforts in raising the 95 percent accuracy figure to a higher one.

Suppose we were incredibly successful at this effort and developed a screening test which had a 100 percent sensitivity. Then, in our numerical example, the number of people who receive positive tests increases from 95 + 999 to 100 + 999 = 1099. The conditional probability that a person has the disease given that the test is positive now becomes

100/1099 = .091

which is just over 9 percent. We have not improved the probability very much.

We've just seen that efforts to improve a screening procedure which already has a high sensitivity by increasing that sensitivity are not likely to increase dramatically the chances that a positively tested individual has the disease.

The other problem with boosting sensitivity is that it often increases the probability of false positives. The screening technique may be based on a "threshold " value for some substance. If the measure is above the threshold value, then we call the test result positive. One way to improve the sensitivity of the test is to lower the threshold value. Then more people who actually have the disease will receive positive tests. Unfortunately, so will more people who don't have the disease.

To illustrate the drawbacks of this approach, suppose the sensitivity is raised to 100 percent at the cost of raising the percentage of positive tests among those free of the disease from 1 percent to 2 percent. Then of the 100,000 people tested, all 100 of the disease victims are identified as positive. Of the 99,900 disease-free people, 1998 will test as positive. The probability of having the disease given a positive test is

100/(100 + 1998) = 100/2098 = .05

which is worse than the 8 percent probability we had originally.

### 8.4.2 Specificity

As a second attempt to improve the situation, we might bend our efforts to reduce the error rate of false positives. The specificity of a screen test is a measure of how well the test performs in this aspect. The specificity is the

probability of a disease-free person having a negative test. The more specific a test is, the smaller is the proportion of false positives.

In our original numerical example, the specificity was 99 percent so that the probability of a positive test for a disease-free person was .01. Suppose we could reduce this probability by a factor of three. Then the number of people with false positives would drop from 999 to 333. The probability that a person has the disease given a positive test would then be

$$95/(95 + 333) = .22$$

which is a significant improvement.

In many screening procedures, increased specificity comes only at the cost of decreased sensitivity. If we raise the threshold for a positive test, then fewer people without the disease will test positive. Hence the specificity will go up. But then more people who actually have the disease will also test negative and we will fail to diagnose their illness. In our continuing example, suppose that effort which reduced the number of false positives by a factor of three also lowers the sensitivity from 95 percent to 90 percent of the population. Then the proportion of those with positive tests who have the disease becomes

$$90/(90 + 333) = .21$$

which is not very different from the 22 percent figure we obtained above.

If our primary goal is to raise the probability that a positively tested person actually has the disease, then our analysis so far shows that we should concentrate our efforts on improving the specificity of the screening procedure even at the cost of lowering the test's sensitivity. We should realize, however, that the pursuit of this primary goal may mean a greater failure to detect the disease in those who have it at a early stage where effective treatment may be possible.

### 8.4.3 Incidence

There is a third important factor that enters into these calculations. This is the relative incidence of the disease among the population being tested. In our example, we assumed that about one person per thousand actually had the disease. If the actual proportion is as high as one in a hundred, then the probability that a person with a positive test has the disease will change. Of the 100,000 people in the study, about 1000 will have the disease. Among these 1000 we expect that 950 (95 percent) will test positive. There will be about 99,000 people without the disease. One percent of them, or 990 will

test positive. Thus there will be about 1000 + 990 = 1990 positive tests. The proportion of disease carriers among the positive tests will be

1000/1990 = .503, or just over one-half.

If our disease is present in one out of every ten people, then the test will be positive for about 9500 of the 10,000 people with the disease. The test will also be positive for about 900 of the 90,000 who don't have the disease. In this case, the probability that a positively tested person has the disease jumps to

9500/(9500 + 900) = .91

so that less than 10 percent of those with positive tests are actually free of the disease.

These last examples indicate that certain aspects of medical screening may be most effective when applied to a population which is believed to have a much higher rate of incidence of the targeted disease than the general population. In March, 1986 U.S. health officials recommended that Americans at "high risk" of contracting AIDS should undergo periodic blood tests to determine if they have become infected with the virus. Those seen as having an increased risk of infection include:

— Homosexual and bisexual men.
— Present or past intravenous drug users.
— People with signs of symptoms compatible with AIDS.
— People born in Haiti and certain central African nations.
— Male or female prostitutes.
— Sexual partners of infected people or of people at increased risk.
— Hemophiliacs who have received blood-clotting factor products.
— New-born children of infected mothers.

### *8.5 A More Formal Solution*

In this section, we use the basic probability rules to derive an expression for the conditional probability that a person has AIDS given that the test result was positive.

Suppose

Pr(AIDS) = Probability that randomly chosen person has AIDS = a

so that

Pr(No AIDS) = Pr(AIDS*) = 1 - Pr(AIDS) = 1 - a

and suppose

Pr(Positive | AIDS) = Probability that test indicates presence of AIDS given that person actually has AIDS ("true positive") = p

so that

Pr(Negative | AIDS) = Probability that test fails to show presence of AIDS given that person actually has AIDS ("false negative") = 1 - p.

Finally, suppose

Pr(Positive | AIDS*) = Probability that test indicates presence of AIDS given that person does not have AIDS ("false positive") = q

so that

Pr(Negative | AIDS*) = Probability that test fails to show presence of AIDS given that person does not have AIDS ("true negative") = 1 - q.

Thus Pr(Positive Test) = Pr(Positive,AIDS) + Pr(Positive,AIDS*)
= Pr(Positive | AIDS) * Pr(AIDS) +
Pr(Positive | AIDS*) * Pr(AIDS*)
= pa + q(1-a)

We want Pr(AIDS | Positive).

Now, Pr(Positive,AIDS) = Pr(AIDS,Positive)
= Pr(AIDS | Positive) Pr(Positive)

Hence

$$\Pr(\text{Positive} \mid \text{AIDS}) = \frac{\Pr(\text{AIDS, Positive})}{\Pr(\text{Positive})} = \frac{pa}{pa + q(1 - a)}$$

### *8.6 Lie Detector Tests*

If you are unjustly accused by the police of a crime you didn't commit, should you volunteer to take a "lie detector" test to prove your innocence?

The polygraph is a psychophysiological device that records physiological variables such as heart and respiration rates, blood pressure and galvanic skin response, all of which are under autonomic control. Advocates of poly-

graph use claim that when these variables are measured while a subject is interrogated, then there is a characteristic pattern of responses made by lies that can be detected by a trained examiner. In a recent study (" Predictive Power of the Polygraph: Can The 'Lie Detector' Really Detect Lies?", *The Lancet*, March 8, 1986), three physicians examined a number of experimental studies about polygraphs. They found an average sensitivity of .76 with a low mark of .64 and a high one of .82. The average value of specificity in these studies was .63, ranging from a low of .5 to a high of .82.

Using the average values of .76 for sensitivity and .63 for specificity, they constructed a table showing the expected results if a polygraph test was administered to a population of 1000 people, five percent of whom were actually lying:

| Polygraph result | Liars | Truth tellers | Total |
|---|---|---|---|
| Positive (read as lying) | 38 true positives | 351 false positives | 389 total positives |
| Negative (read as truthful) | 12 false negatives | 599 true negatives | 611 total negatives |
| Total | 50 | 950 | 1000 |

From this table, we can easily calculate the predictive value of a positive test. Recall that this is the probability that a person with a positive polygraph test is lying. This would be

$$\frac{\text{True Positives}}{\text{Total Positives}} = \frac{38}{389} = .098$$

or just under 10 percent.

# Chapter 9

**The Stolen Cars Problem**

## *9.1 The Questions*

In our problem, thieves make off with 20 percent of all the cars in the neighborhood each year. We can easily pose three questions of interest to us:

1. What are the chances that my car will be stolen this year?

2. How many years can I expect to go by before thieves get my car?

3. How long will it take before everybody in the neighborhood suffers from auto theft?

## *9.2 Some Simplifying Assumptions*

We'll see that it's not difficult to answer these questions if we make a few simplifying assumptions about the situation. We'll assume that each person's car is as likely to be stolen as any other person's car. We'll also assume that successive car thefts are independent of each other. We're asserting here that you and your next-door neighbor each have the same chance (1 in 5) of being the next victim of automobile theft. We're also saying that the chances that the next car stolen belongs to a particular person are not dependent on the previous history of car thefts.

In the real world, these assumptions are probably not strictly valid. Whether a thief decides to go after your car rather than your neighbor's may depend on such factors as the age, make, or model of the car, whether the car is on the street or in a garage, whether the owner is home or not, etc. Also, if your car is stolen one year, you may take extra security measures in subsequent years that will deter thieves. The likelihood of your being a victim again may go down. On the other hand, since the thieves were successful against you last time, they may be more likely to return to the "scene of the crime" and try again.

## *9.3 The Answers*

The answers to our first questions are easily obtained from our previous work. For our basic model, we're asserting that there is a probability of .2 that any particular car will be stolen in a given year. Thus, you have a 20 percent chance of having your automobile stolen this year and an 80 percent chance of completing a full year with uninterrupted possession of your car.

If the thefts are independent of each other, then the probability of two successive years without suffering a theft is

$(.8)(.8) = .64$

so there is a probability of .36 of having at least one car stolen in two years.

The likelihood of going three years in a row without a car theft is

$(.8)(.8)(.8) = .512$

or a little more than one-half.

From our section on expected value, we see that on average, you should expect five years to go by before your car is stolen. You should compute the probability that your car has not been stolen during a five-year period; the answer is $.8^5 = .32768$.

The one problem we have not yet solved is how long it should take on the average before everyone in the neighborhood experiences a stolen car.

A first approach to this problem is to do some simulations. You could simply wait and observe the car thefts in the neighborhood for several years. You might even join (or organize) a car theft ring and keep careful statistics on your operations. Such methods of collecting experimental insight into the problem might have severe consequences for you.

A cheaper *physical simulation* is available. Take an ordinary six-sided die (that's one of the members of a pair of dice). Roll it repeatedly until each of the numbers one through five has been observed. Keep track of the number of rolls until you've obtained all 5 numbers; simply ignore any of the rolls which results in a six.

In this approach and our subsequent analysis, we'll assume for simplicity that there are exactly five different families in our neighborhood each with one automobile and that one car theft occurs each year. Also to simplify our discussion, we'll assume that each stolen car is recovered or replaced by an identical model.

A third alternative is to write a computer program to simulate the process. Appendix A contains a simple BASIC program that does this. When this program was run 1000 times, we found that the average number of thefts that occurred before all five cars were stolen was 11.59.

The following table gives a more detailed report. It shows how many times we observed 5 thefts, 6 thefts, etc.

| Thefts | Number of Times |
|---|---|
| 5 | 39 |
| 6 | 76 |
| 7 | 116 |
| 8 | 103 |
| 9 | 93 |
| 10 | 99 |
| 11 | 71 |
| 12 | 63 |
| 13 | 56 |
| 14 | 48 |
| 15 | 45 |
| 16 | 33 |
| 17 | 33 |
| 18 | 22 |
| 19 | 17 |
| 20 | 18 |
| 21 | 14 |
| 22 | 13 |
| 23 | 7 |
| 24 | 9 |
| 25 | 5 |
| 26 | 2 |
| 27 | 2 |
| 28 | 5 |
| 29 | 2 |
| 30 | 0 |
| 31 | 1 |
| 32 | 2 |
| 33 | 2 |
| 34 | 1 |
| 35 | 0 |
| 36 | 0 |
| 37 | 0 |
| 38 | 2 |
| 39 | 1 |

Note that in just under 4 percent of the trials (39 of 1000), we found all 5 cars stolen in the first 5 thefts. In the most extreme cases, we had 39 thefts before getting a complete set of cars. In 90 percent of the trials, at most 19

before getting a complete set of cars. In 90 percent of the trials, at most 19 thefts were required. In only 1 percent of the trials did we need to look at more 28 thefts.

Let's see if we can attack the problem analytically. It is easy to find the expected number of thefts to obtain the first car. It will be one.

Now let's jump ahead a little bit. Suppose that we have two different cars stolen already. We wish to determine the expected number of thefts to get a third car different from the first two. There are 3 cars we haven't seen yet. Let's call it a success if a theft results in one of these three cars. For any randomly chosen theft, the probability of a success is 3/5. As we saw earlier in our section on Expected Value, the expected number of trials for a success is 5/3.

What we have just done in finding the expected number of thefts to find the third car can be imitated for each of the other stages as well. If we have three different cars and are searching for the fourth, then either of the two remaining cars will do. There is a probability of 2/5 that any single theft will yield a new car. The expected number of thefts for the fourth car is 5/2.

This reasoning holds equally well for other numbers of cars. We summarize the results in the following table :

| Car | Probability of Success | Expected Number of Thefts To Get This Car |
|---|---|---|
| First | 1 | 1 |
| Second | 4/5 | 5/4 |
| Third | 3/5 | 5/3 |
| Fourth | 2/5 | 5/2 |
| Fifth | 1/5 | 5 |

Putting this information together with the definition of expected value, we see that the expected number of thefts to get all 5 cars is

$$1 + (5/4) + (5/3) + (5/2) + 5 = 11\ 5/12 = 11.417$$

Since there is one theft per year for any group of five cars, we conclude that on the average we would expect a period of about 11 and a half years before everyone is the victim of a automobile theft.

# Chapter 10

**Reliability of Warning Devices**

Our newspapers and television broadcasts often present us with graphic images of the consequences of technological failures: the Challenger explosion, the Three Mile Island nuclear plant accident, the release of poisonous gasses from a chemical plant in India, the midair collision of two passenger airplanes, to name but a few.

## *10.1 The Promise and Perils of Warning Devices*

In an effort to prevent such disasters, engineers often install warning devices inside of complex technological artifacts. The warning devices should alert the human workers to take emergency action.

Consider, for example a nuclear reactor plant generating electricity for home and commercial uses. A warning device should give a clear alarm if it detects an unacceptably high level of radiation. The human response would then be to shut down the plant and begin a careful inspection of the reactor.

The United States and the Soviet Union each maintain systems to warn if a massive missile strike has been launched by the "other side." Some military experts have argued that there is insufficient time to alert the American President or the Soviet Premier and wait for a decision; the only "reasonable" response is an automatic launching of a retaliatory strike. Thus they call for an automation of the system; when the warning bell rings, the system responds with human intervention.

In both of these cases — and a host of similarly structured situations— there are serious consequences if the warning device does not function properly. The warning device could fail by

1. Not sounding its warning when there is excessive radiation or there has been an enemy missile strike; or

2. Falsely issuing a warning when there is no danger.

Here are some other examples of real world situations in which warning devices play an important role :

— Home smoke detectors;
— Devices on board commercial aircraft that alert pilots if another plane is on a collision course;

— National Weather Service declarations of hurricane or tornado alerts and warnings;
— Radar detectors on board trains which warn of obstacles in the path of these high speed vehicles.

As in the case of our AIDS screening problem, we are faced with a mechanism that may give False Positives or False Negatives. One solution calls for installing several copies of the warning device in the system. If the social and economic costs of a malfunctioning warning device are high and the cost of multiple warning devices is low, then using several devices could be effective if we can show that this duplication increases reliability.

### *10.2 The Two Device Situation*

Let's begin with a simple example. Suppose that a single warning device functions properly 90 percent of the time. About 90 out of 100 times when there is a true danger present, the device will alert us. Similarly, in only about 10 of 100 times when there is no true danger will it falsely signal. We install two copies, A and B, of this device. We'll assume that the two devices operate independently of each other. Then, in the case of true danger, the probability that both A and B alert us is (.9)(.9) = .81. The table below summarizes the possible cases:

| | True Danger | No Real Danger |
|---|---|---|
| A warns, B warns | (.9) (.9) = .81 | (.1) (.1) = .01 |
| A warns, B silent | (.9) (.1) = .09 | (.1)(.9) = .09 |
| A silent, B warns | (.1) (.9) = .09 | (.9) (.1) = .09 |
| A silent, B silent | (.1) (.1) = .01 | (.9)(.9) = .81 |

What should be our policy? There are basically two policy decision rules we could follow:

1. Act as if there is a real danger only when both A and B sound the signal.

If there is a true danger, we will act appropriately 81 percent of the time. If there is no real danger, then we will take inappropriate action only 1 percent of the time.

2. Act as if there is a real danger if at least one of the devices sounds the signal.

If there is a true danger, the probability that at least one of the devices works correctly is .99 (that is, .81 + .09 + .09). If there is no real danger, then we will take unnecessary actions 19 percent of the time.

Notice that the first policy decreases the likelihood of correctly identifying a real danger while increasing the chances of correctly ignoring a false danger signal.

The second policy has the reverse effect. It makes it more likely that we will recognize danger when it really is there, but it also boosts the chances that we will assume a dangerous situation when there actually isn't one.

### *10.3 The Three Device Situation*

Going from one warning device to two hasn't solved our problem for us. But consider what would happen if we installed three devices (A, B and C), each with the same individual reliability and operating independently of the others.

| | True Danger | No Real Danger |
|---|---|---|
| A warns, B warns, C warns | (.9) (.9) (.9) = .729 | .001 |
| A warns, B warns, C silent | (.9) (.9) (.1) = .081 | .009 |
| A warns, B silent, C warns | (.9) (.1) (.9) = .081 | .009 |
| A warns, B silent, C silent | (.9) (.1) (.1) = .009 | .081 |
| A silent, B warns, C warns | (.1) (.9) (.9) = .081 | .009 |
| A silent, B warns, C silent | (.1) (.9) (.1) = .009 | .081 |
| A silent, B silent, C warns | (.1) (.1) (.9) = .009 | .081 |
| A silent, B silent, C silent | (.1) (.1) (.1) = .001 | .729 |

Here we can formulate three decision-making policies:

1. Assume there is danger only if all three devices sound the alarm. Where this is a real danger, this will happen with probability .729. In the cases where there is no real danger, we will falsely believe there is danger with a probability of only .001.

2. Assume there is danger if at least one device sounds. Under this decision policy, we will make the right decision in 999 of a 1000 cases when there is a true danger, but the probability of our assuming there is a risk when no danger is present jumps to .271.

3. Assume there is a danger if a majority of the devices (2 or 3) sound a warning.

When there is a true danger, this policy recognizes it with probability .971 (that is, .729 + .081 + .081 + .081). When no danger is present, we falsely assume it is with probability .028.

The majority rule policy is clearly the best one. It improves upon the chances of acting correctly in both the True Danger and False Danger situations. Compared with having a single warning device, using a two-out-of-three-or-better boosts the chances of making the correct judgment in a true danger situation from .9 to .971. It also decreases the probability of making the wrong judgment when no danger is present from .1 to .028.

# Appendix A

## Computer Programs

We collect here BASIC programs that were used to perform some of the calculations and simulations discussed in the body of the module.

These programs should run on IBM and IBM-compatible microcomputers running BASICA or on DEC Rainbows and other MS-DOS machines using GW-BASIC. Modifications to other dialects of BASIC should be minor.

### A BASIC program to compute N Choose K

```
100 ' N CHOOSE K
110 ' Program computes the value of N CHOOSE K
120 ' which is the fraction N!/[K!  (N-K)!]
130 ' The program uses the fact that the value is equal
140 ' to (N/K)* (N-1) /(K-1)* (N-2) /(K-2)*... (N-K+1)/1
150 '
200 INPUT "Enter N and K"; N,K
210 LET VALUE = 1
220 FOR J = 1 TO K
230    LET VALUE = VALUE * (N - K + J)/J
240 NEXT J
250 PRINT "VALUE OF";N;" CHOOSE";K;"IS"; VALUE
300 END
```

### The Witches Example

```
100 ' WITCHES
110 ' Program simulates the random selection of
120 ' 16 objects from a population of 89 of which
130 ' 21 are labelled MALE and 68 FEMALE
140 '
150 DIM P(89)
160 FOR I = 1 TO 89
170    LET P(I) = 0
180 NEXT I
190 '
200 RANDOMIZE TIMER
210 LET MALE = 0
220 LET FEMALE = 0
225 LET CHOICE = 0
230 WHILE CHOICE < 16
240   GUILTY = 1 + INT(89*RND)
245   IF P(GUILTY) = 1 THEN GOTO 260
250   IF GUILTY <= 21 THEN LET MALE = MALE + 1
          ELSE  LET FEMALE = FEMALE + 1
255   P(GUILTY) = 1
260 WEND
270 '
300 PRINT "There were ";MALE;" men and ";FEMALE;" women selected"
310 END
```

### Stolen Cars

```
100 RANDOMIZE TIMER
200 DIM A(5) ' indicates which cars have been stolen;
210 FOR I = 1 TO 5 ' a 1 means car has been stolen.
220    LET A(I) = 0
230 NEXT I
300 LET THEFTS = 0 ' counts total number of car thefts
310 LET CARS = 0 ' counts number of different cars stolen
320 WHILE CARS < 5
330  LET STICKER = 1 + INT(5*RND) :
335  LET THEFTS  = THEFTS + 1
340  IF A(STICKER) = 0
     THEN CARS = CARS + 1 : A(STICKER) = 1
350 WEND
360 PRINT "Number of Thefts is "; THEFTS
```

# Appendix B

**Deriving Some Expected Value Formulas**

How many times would you expect to flip a fair coin before it landed "Heads"? We might get a Heads as early as the first flip. In fact, we would anticipate that this would happen about half the time. We might not get a Heads, however, until the second flip. This would happen if the sequence was

Tail, Head

which would occur with probability

$(1/2)\ (1/2) = (1/2)^2$

Now the initial Heads might not appear until the third flip. For this to be the case, the sequence must

Tail, Tail, Head

In the general case, the first Head might not appear until the kth flip. For this to happen, our sequence of flips must be (k-l) Tails followed by a Head:

Tail, Tail, ..., Tail, Head
(k - 1) times

This occurs with probability

$(1/2)\ (1/2)\ldots\ (1/2)\ (1/2) = (1/2)^k$
(k - 1) times

We can now write the formula for the Expected Value of the number of flips on which the first Head appears:

$$V = (1)\left(\frac{1}{2}\right) + (2)\left(\frac{1}{2}\right)^2 + (3)\left(\frac{1}{2}\right)^3 + (4)\left(\frac{1}{2}\right)^4 + \ldots$$

This representation of the Expected Value V is in terms of a sum of infinitely many terms. We would like to find a "closed form" or more compact

description of V.

In order to do this, we'll need to derive two more general results. These will not only give us an alternative, simpler expression for V; they will also be very useful in a number of related problems.

First, consider the "sum"

(*) $S = 1 + x + x^2 + x^3 + x^4 + x^5 + \ldots$

where each term is x multiplied by the previous term. Now let's multiply each side of this equation by x:

$xS = x(1 + x + x^2 + x^3 + x^4 + x^5 + \ldots)$

or

(**) $xS = x + x^2 + x^3 + x^4 + x^5 + \ldots$

Now let's line up equation (**) below equation (*) and subtract the bottom one from the top:

(*) $S = 1 + x + x^2 + x^3 + x^4 + \ldots$
(**) $xS = x + x^2 + x^3 + x^4 + \ldots$

$S - xS = 1$ (subtracting like powers of x on the right side)

$(1 - x)S = 1$ (factoring a common S from both terms on left)

or, finally,

$S = 1/(1-x)$ (dividing both sides by 1-x )

We've taken some mathematical liberties in this derivation. You should be reassured that everything can be justified with the utmost rigor provided x is any number strictly between -1 and +1. The general result can be summarized:

If x is any number between -1 and 1,
then $1 + x + x^2 + x^3 + \ldots + x^k + \ldots = 1/(1-x)$

For our second result, we want to examine a more complicated sum:

$$V = x + 2x^2 + 3x^3 + 4x^4 + 5x^5 + \ldots + kx^k + \ldots$$

As before, multiply both sides of this equation by x and then subtract:

$$\begin{array}{rcl} V & = & x + 2x^2 + 3x^3 + 4x^4 + 5x^5 + \ldots \\ xV & = & x^2 + 2x^3 + 3x^4 + 4x^5 + \ldots \\ \hline V - xV & = & x + x^2 + x^3 + x^4 + x^5 + \ldots \end{array}$$

Note that the columns don't quite "cancel out" as before, but the resulting sum on the right hand side has a familiar form. Factoring out an x from each term, we have

$$(1-x)V = x(1 + x + x^2 + x^3 + x^4 + x^5 + \ldots)$$

so we can apply our previous result to obtain

$$(1-x)V = x\,[1/(1-x)] = x/(1-x)$$

and dividing both sides by l-x gives

$$V = x/(1-x)^2$$

We summarize this result in the statement:

If x is any number between -1 and 1, then

$$x + 2x^2 + 3x^3 + 4x^4 + 5x^5 + \ldots = x/(1-x)^2$$

If we substitute x = 1/2 into this last equation, we see that the left hand side is exactly the expected value of the number of flips necessary to obtain a Head, while the right hand side is

$$\frac{x}{(1-x)^2} = \frac{1/2}{(1/2)^2} = \frac{1}{1/2} = 2$$

We conclude from this that, on the average, we would have to flip a fair coin twice before we saw a Head appear. We can make a slightly more general conclusion. If an outcome of an experiment has probability 1/2 of occurring, then on average, we should expect to repeat the experiment twice before the event occurs. For example, if we roll a fair die, then the probability of obtaining an odd result (1,3 or 5) is 1/2, so we would expect to roll the die twice on the average in order to get an odd number.

## Suggested Exercises and Projects

1. Suppose automobile license plates consist either of a single letter followed by three digits or a pair of letters followed by two digits. How many different license plates are possible?

2. In a round-robin basketball tournament each of 16 teams will play each other once. How many games will be played? In a single elimination tournament, like the NCAA competition, a team can't advance to a subsequent round unless it wins the current game. How many games are played in such a tournament?

3. A small New England college divides its curriculum into four areas. Certain courses in each division have been designated as "Foundation Courses." The divisions and the respective number of Foundations courses are

| | |
|---|---|
| Humanities | 20 |
| Foreign Languages | 16 |
| Social Sciences | 8 |
| Natural Sciences | 17 |

Each student must take a Foundations course in at least three of the four divisions. In how many different ways can a student satisfy this requirement?

4. You decide to invest a large inheritance in a portfolio of eight stocks and four bonds. Your broker recommends 12 different stocks and 7 bonds as good investments. In how many ways can you create the portfolio?

5. Find the probability of obtaining 3 Heads and 2 Tails in five tosses of a fair coin.

6. Find the probability of getting a sum larger than 7 when rolling a pair of fair dice.

7. A monkey sits at a typewriter with keys for the 26 letters of the alphabet and a space bar. If the monkey hits 3 keys at random what is the probability that it spells APE? What is the probability that the monkey will strike the pattern Consonant-Vowel-Consonant?

8. If you choose 3 people at random and ask about the month in which they were born, what is the probability that at least two of them were born in

the same month? Suppose you choose 7 people; what is the probability? What is the probability if you ask 13 people?

9. A poll of 40 freshman and 45 sophomores found 21 sophomores preferred diet soda and 47 people favored regular soda. What is the probability that a person randomly chosen from the poll prefers diet soda? What is the probability that a randomly chosen person is a freshman who favors regular soda? Suppose a person is chosen at random and we learn that she prefers diet soda; what's the probability that she is a freshman?

10. Two metal detectors designed for airport security checks are on the market. The first costs $20,000 and will detect weapons with a probability of .999. The second has a lower detection probability of .99, but costs only $10,000. It's possible to install two of the cheaper devices in a way in which they will act independently of each other; if this is done, then a passenger must step through each of the devices in turn. Is it better to buy the two $10,000 detectors or one $20,000 device?

11. Fly-By-Night Computer Sales offers bargain prices on computer systems. Unfortunately, not all the equipment they ship works. About 20 percent of the printers, 15 percent of the central processing units, and 12 percent of the monitors that consumers receive are faulty. You're thinking about purchasing a system from this company. Determine the probability that

(a) the monitor and printer will be defective
(b) all three components won't work
(c) all three components will work
(d) at least one component will fail

12. If a coin is flipped 8 times, find the expected number of heads assuming

(a) the coin is a fair one;
(b) the coin is weighted so that it comes up heads with probability 3/5.

13. My automobile insurance company charges $500 a year for an accident policy. It claims that about 8 percent of its clients file a claim each year and the average claim is $1200. What is the company's expected gain per policy?

14. In the game of "craps" a player rolls a pair of dice. If the sum of the numbers shown is 7 or 11, she wins. If it is 2,3 or 12, she loses. If it is any

other sum, she must continue rolling the dice until she either repeats the same sum (in which case she wins) or she rolls a 7 (in which case she loses). Suppose the outcome of each round is a win or a loss of $5. What is the probability that our player wins a round? What is the expected value of shooting craps?

15. A roulette wheel has the numbers 0,1,2,...,36 marked on 37 equally spaced slots. The numbers from 1 to 36 are evenly divided between red and black. A player may bet on either color. If a player bets on red and a red number turns up after the wheel is spun, he receives twice his stake. If a black number turns up, he loses his stake. If 0 turns up, then the wheel is spun again until it stops on a red or a black. If this is red, the player receives only his original stake and if it is black, he loses his stake. If a player bets $1 on red with each spin, what is his expected value of winning?

16. (Expected value considerations applied to decision-making) A construction firm is considering bidding on a contract to build one of two new schools. The firm may submit a bid to construct a high school. It estimates that there would be a $50,000 profit on this project, but that it would cost $1,000 to prepare the plans that must be submitted with the bid. The other possibility is a bid on an elementary school. The firm has built several elementary schools in the recent past and estimates that the preparation costs for a bid would be only $500 while the potential profits is $40,000. In estimating profit, all costs—including that of the bid—have been considered.

The construction company only has enough time to prepare one bid. Past experience leads the company to estimate that it has one chance in five of winning the high school contract and one chance in four for the elementary school. The firm's goal is to maximize expected profit. Which bid should the company prepare and submit?

17. There were 60 finalists for a national fellowship program awarding 35 prizes. Of the finalists, 15 were women and 45 were men. Only 5 of the women received prizes, while 30 of the men were winners. Is there sufficient evidence here to conclude that gender played a role in the selection of the winners?

18. During an eleven year period, 870 people from one Texas county were summoned as potential grand jurors. Mexican-Americans made up 79 percent of the county's population, but only 339 Mexican-Americans (39 percent) were selected for grand jury duty. What calculations would you do to determine if the small number of Mexican-Americans was due to chance? The United States Supreme Court dealt with this problem in the 1977 case *Castaneda vs. Partida* (430 U.S. 482, 51 L. Ed. 2d 498). The Court accepted arguments based on probability models that grand jury selection was not free of discrimination against Mexican-Americans.

19. If the sensitivity of a screening process is .98 and the specificity is .92, and if the process is used to detect a condition present in 5 percent of the population, determine the probability of having the condition given a positive test.

20. A multiple-choice exam has four suggested answers to each question, only one of which is correct. A student who has done her homework is certain to identify the correct answer. If a student skips her homework, then she chooses an answer at random. Suppose that two-thirds of the class has done the homework. While grading, the teacher observes that Jennifer has the right answer to the first problem. What is the probability that she did the homework?

21. A 1983 study of United States Army personnel on active duty yielded the following figures:

| | Men | Women |
|---|---|---|
| Commissioned Officers | 80,091 | 8,921 |
| Enlisted Personnel | 606,956 | 67,077 |

Using this data, determine:

(a) Total number of men, women, commissioned officers, and enlisted personnel.
(b) The probability that a randomly chosen Army member is
  (i) a women
  (ii) a male commissioned officer
  (iii) an enlisted person
(c) The probability that a randomly chosen female Army member is an officer.
(d) The probability that a randomly chosen officer is a woman.

22. The Internal Revenue Service uses computers to identify suspicious tax returns that warrant further auditing. Approximately 2 percent of all returns are spotted by the computer for auditing. Of those returns which are audited, about 85 percent turn out to be ones in which the taxpayer owes the IRS more money. Of the returns that are not audited, experts estimate that

perhaps one in five are cases where taxpayers have underpaid. Suppose a tax return is randomly selected. What is the probability that it represents a case of underpayment? If it is an underpayment, what is the probability that the computer will spot it for an audit?

23. Suppose you move to a safer neighborhood where the incidence of auto theft drops from 20 percent to 10 percent. Determine the expected number of thefts before everyone has a car stolen.

24. National safety experts estimate that one in every four Americans is injured each year in a household accident. Calculate the expected number of years before every family suffers an accident.

25. Your favorite fast food restaurant CRAMPS has announced a new giveaway. Inside each of its hamburgers is a plastic letter: C or R or A or M or P or S. The first person who collects all six letters will win a prize of $1,000. If equal numbers of each of the letters are inserted at random into the sandwiches, how many hamburgers would you expect to buy before obtaining all six letters? If a substantial number of people each purchase this many burgers and no one has discovered a full set of letters, what conclusions can you draw about the distribution of the letters?

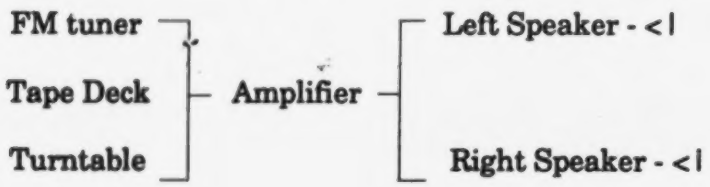
26. Analyze the two- and three- device situations if a single warning device functions properly 95 percent of the time.

27. A warning device will sound its alarm with probability .99 when there is danger present but will falsely ring 3 percent of the time when there's no danger. How does the reliability change if we adopt a majority-rule strategy based on three such devices?

28. Returning to our original situation with a single device operating correctly 90 percent of the time, determine the improvement in reliability if we install 5 of these devices and make our decision using a majority rule.

29. Is majority rule the optimal strategy to follow if we have five identical devices?

30. A typical student stereo system is arranged as follows:



The way the system has been set up, we will be able to hear some music provided

— At least one of the initial devices (FM tuner, Tape Deck, Turntable) works; and

— The amplifier works, and

— At least one of the speakers works

If both speakers work, then we can produce stereophonic sound. Each of the six components has its own reliability; that is, the probability that it will function. The reliabilities are:

| Component | Reliability |
|---|---|
| FM tuner | .98 |
| Tape Deck | .993 |
| Turntable | .95 |
| Left Speaker | .97 |
| Right Speaker | .97 |

Determine the probability that

(a) all components work
(b) we can hear some kind of music out of the system
(c) we can hear some stereo music
(d) the system fails
(e) The tape deck, amplifier and left speaker work
(f) only the tape deck, amplifier and left speaker work

For further information about the Council of Independent Colleges and its project on Technology and the Liberal Arts, write:

CIC
Technology Project
One Dupont Circle, Suite 320
Washington, D.C. 20036

Additional copies of this module can be purchased from the CIC Publications Department at the address above.